A Novel Approach to Chronic Insomnia Prediction Using Machine Learning Algorithms

Md. Lutfor Rahman Abir
Department of Electrical and Computer Engineering
North South University, Bashundhara, Dhaka-1229, Bangladesh

Md. Mashurul Haque Urfi
Department of Electrical and Computer Engineering
North South University, Bashundhara, Dhaka-1229, Bangladesh

Sultan Mahmud Sagor
Department of Electrical and Computer Engineering
North South University, Bashundhara, Dhaka-1229, Bangladesh

Rahima Khatun
Department of Electrical and Computer Engineering
North South University, Bashundhara, Dhaka-1229, Bangladesh

Mohammad Monirujjaman Khan
Department of Electrical and Computer Engineering
North South University, Bashundhara, Dhaka-1229, Bangladesh

Corresponding Author: Mohammad Monirujjaman Khan

Email: monirujjaman.khan@northsouth.edu

***Abstract*—Over the past ten years, the power of technology has profoundly changed the globe. As a result, many changes are occurring in people's lives that have an impact on their health in both positive and negative ways. Overuse of radio-wave devices, inactivity, and virtual participation are all contributing factors to a variety of health issues, including insomnia. The illness is also referred to as sleeplessness. This may can place independently or because of another problem. Chronic (prolonged) insomnia can gravely damage a person's brain, which could result in a lifelong illness. However, several medical tests can detect the presence of insomnia based on numerous internal sleep characteristics. However, this approach is both time- and money-consuming. Many developing countries also lack pricey testing equipment. To close this gap, we have chosen to develop an advanced machine learning-based model that can forecast chronic insomnia. The goal of this study is to look at depressive symptoms using machine learning to increase efficiency and accuracy. It has an excellent union of feature selection approaches and classifiers by studying the combinations of the most prominent feature selection techniques and classifiers, such as K-Nearest Neighbors (KNN) Classification, Decision Tree (DT) Classification, Logistic Regression (LR), Random Forest (RF), Ada Boost (AB), Support Vector Machine (SVM) and Gradient Boosting (GB). The Random Forest classifier was found to be more accurate than any other model with 94 percent accuracy. A lot of the models used in this study are more accurate than the models used in previous studies, which means that the models are more trustworthy. Overall, the study showed that machine learning models are effective in predicting human Insomnia and are a useful tool for locating and predicting the related risk variables that affect mental health. The use of machine learning models for youth Insomnia screening and early detection may make it easier to create health prevention and intervention programs that will improve the mental health of young people in low and middle-income nations.**

***Keywords*— Insomnia; Accuracy; Machine learning; Random Forest; Classifier; Data; Mental Health**

I. Introduction

Lack of sleep is not the only symptom of insomnia. This is frequently acknowledged as an extra risk factor for disease. It is typically accompanied by conditions including anxiety, enuresis, sleep apnea, and nightmares [2,3,4]. It is usually treated as a cardiovascular disease presymptom [1]. This disease is one of the major factors in auto accidents, according to study. It is acknowledged as a risk factor for MVA and NMA [5,6,7]. Adults and the elderly are now the groups most impacted by the disorder. It changes depending on the person. Even though it may come and go, chronic sleeplessness can have a major negative impact on one's mental health. If the sleep disturbance is severe enough, the patient might not be able to get even one hour of sleep per week. Because they are unconscious, people usually do not realize that they have insomnia. Some people think that the reason they can't sleep properly during the primary stage is because they are having such a miserable time. They take sleeping drugs while they wait for regular, long-lasting sleep. Unaware people never even get a checkup at the doctor. They are at odds with such grave conditions because they treat this casually. People frequently misunderstand the causes of sleeplessness. They are unable to differentiate between insomnia and sleep problems. Not getting enough sleep or simply having difficulties falling asleep do not cause insomnia. Insomnia keeps some parameters in place. Doctors frequently diagnose this condition based on the patient's symptoms. They employ a questionnaire for this [8, 9]. A true insomniac (chronic) patient will typically exhibit the symptoms outlined below [10]. The patient is very unhappy with the quality of his sleep; he normally only gets 4-5 hours per night, whereas most people only get between 2 and 3 hours. Also possible are sleep disruptions (frequent waking up during sleep and difficulty falling back asleep), daytime exhaustion, drowsiness, inability to fall asleep during the day while being awake at night, and dread of remaining awake at night. According to the data, mental health issues dropped by 5% as people aged by one year [8]. Children from lower socioeconomic levels who had mental health issues early in childhood—whether because of both parents and just one of them—were more likely to develop mental illnesses later in life. Over the past ten years, discoveries relating to Insomnia and suicide have increased dramatically [9-10]. The prevalence of adult mental health issues has increased, from 10.7 percent in 1996 to 29.2 percent in 2015, according to the Center for Public Health [11]. The most prevalent kind of mental disease, Insomnia, is a psychiatric ailment that can affect anyone at any age for a variety of causes, such as a decline in self-esteem and social circumstances. The symptoms of Insomnia can have a serious impact on a person's capacity to cope with any situation in daily life that differs greatly from normal mood swings. Insomnia has an impact on both physical and psychological health [12]. Back discomfort, diabetes, and hypertension are linked to it [13]. In addition, families, friends, careers, and other connections are frequently burdened by a mental illness in the form of conflict, marriage dissolution, or homelessness [14]. Therefore, a proactive approach and dedication to Insomnia prevention and treatment are required. Considering its prevalence and severity, Insomnia is one of the most prevalent mental illnesses that receives the fewest diagnoses. Nearly all of the information used in the diagnosis and evaluation of depressive symptoms comes from patients, family members, friends, or carers [15]. However, this kind of article is unreliable because it depends on the reporter's complete honesty. Worldwide, Insomnia-related self-perceived guilt is pervasive and is linked to a lack of motivation to seek professional help [16]. A discussion of Insomnia frequently depends largely on a general practitioner's desire to interact with the patient because patients are reluctant to communicate their depressive symptoms with medical professionals. In comparison to the United States and the majority of other Western nations, Malaysia has a significantly greater prevalence of Insomnia [17]. Insomnia is a serious mental condition that has a profound impact on society and on public health. Suicide can result from Insomnia in the worst-case scenario. Less than half of those who have this emotional difficulty have sought mental health treatment, despite the fact that it is a serious psychological disorder. Numerous factors, including ignorance of the condition, may be at blame. Additionally, researchers found that self-stigma and embarrassment appear to play a bigger role in delaying medical care than genuine prejudice and negative reactions from others [18].

With the use of phonic spectral features, Stolar et al. discovered the advanced spectral roll-off set in improvement. For both males and females, the average classification accuracy for all the characteristics that included the best individual spectrum was 71.4 percent and 70.6 percent, respectively [19]. In order to identify the elderly patients who had previously experienced symptoms of Insomnia, four common classification prototypes Bayes Network, C 4.5 Decision Tree, Support Vector Machine (SVM), and Artificial Neural Network (ANN) were used. Soundariya et al. [20] found that ANN produced the best results. Checking whether Insomnia has been disclosed Tsugawa et al. looked into the user's social media habits. Through tests, it was demonstrated that features gleaned from user activity may predict users' Insomnia with an accuracy of 69 percent [21]. Haque et al. studied the vocalized language, and the 3D face features to determine the severity of the Insomnia. This study compared the integrated Convolutional Neural Network (CNN) model. This model indicated 83.3 percent sensitivity and 82.6 percent specificity [22]. Aldarwish and Ahmed used SVM and Naive Bayes models on the previously processed social media posts. They developed a web application to categorize SNS users that depressed patients and psychiatrists could utilize. There is potential to improve the accuracy of this model through training and developing better models [23]. De Choudhury et al. developed an estimated accuracy that was based on what depressed Twitter users were doing. With the aid of multiple people, they were able to acquire the training data for machine learning. They used SVM to collect Twitter user activity in order to forecast the likelihood that any given user will experience Insomnia. An approximate accuracy of 70% was revealed by experimental results [24]. To recognize depressive symptoms in a person's tweets On the Twitter data set, Shetty et al. used a machine learning classifier. They retrieved Twitter posts from a developer's Twitter account using the Twitter API [25]. The accuracy measured by Cao et al. was 84.21 percent. They used feature selection and an SVM model built on the functional connections of resting-state FMRI to categorize patients with severe Insomnia [26]. When Liao et al. used SVM to categorize patients with severe Insomnia based on resting-state EEG data, they achieved an accuracy of 80% [27].

In this study, we used seven machine learning algorithms, and they were all accurate to better than any previous work. Using the Random Forest classifier, we achieved maximum accuracy of 94%. As a result, we were able to narrow the accuracy gap in our research. Adoption of these seven models has the advantage of enabling comparisons to be made. These comparisons assist us in determining which model is the most accurate. The following is the remainder of the article: The experimental strategy and techniques are described in Section 2, while the findings are analyzed in Section 3 and the conclusion is discussed in Section 4.

II. Experimental Approach and Procedures

This section contains some highly specialized information. It has a lot of information about the dataset, the proposed system, and how the research was done.

A. Proposed System

Figure 1 depicts the block diagram of the proposed system.

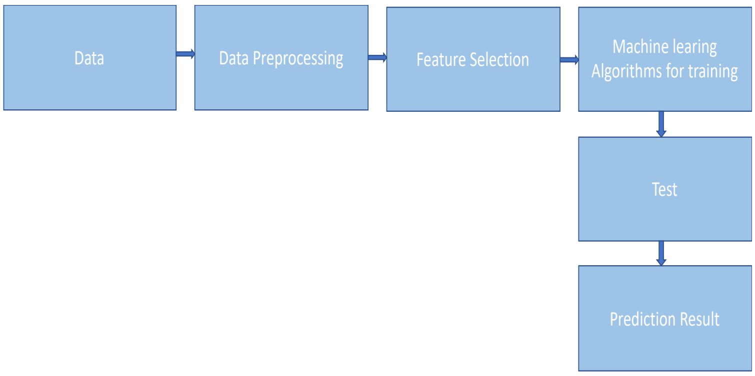

Figure 1: Block Diagram of the Proposed System

To construct a model, the data must be available once the preparation stage has been completed. Data preprocessing is a critical stage in machine learning that improves the quality of the data to encourage the extraction of valuable insights from it. Preparing (cleaning and arranging) raw data in order to make it acceptable for creating and training machine learning models is known as data preprocessing in machine learning. Real-world data frequently lacks attribute values or trends and is frequently inconsistent, erroneous (containing errors or outliers), incomplete, and inconsistent. Data preparation comes into play in this situation because it helps to clean, format, and organize the raw data, making it ready for use by machine learning models. In this research, missing values were dropped during the preprocessing step. After that, we selected the important features to train the machine learning algorithms. Random forest, Logistic

Regression, K-nearest neighbor, Decision Tree, Gradient Boosting, Support Vector Machine, and AdaBoost algorithms were used in this study.

.

B. Dataset

We use the Insomnia dataset [28] for our experimental approaches. In the analysis of Insomnia, the dataset is used. The information was compiled as part of a study on the living circumstances of those who reside in rural areas. In this dataset, there were 1521 rows and 23 columns. In this dataset, the 'depressed' column is the target column. In this column, 1174 data are healthy people's data and 327 data are depressive people's data.

C. Data Preprcessing

Data preprocessing is the process of preparing raw data in such a manner that it may be used directly by a machine learning or deep learning model. It's the most crucial and first step in creating machine learning models. In real-world data, noise, missing values, null values, and data in an unsuitable format for ML models are rather prevalent. Cleaning and preparing data for inclusion in the model are examples of data preparation techniques that may help enhance the accuracy and efficiency of a machine learning model. Our machine learning model may suffer significantly if our dataset contains missing data. As a consequence, it's critical to handle the dataset's missing values and fill in the null values. The dataset is verified for null and missing values. In the no lasting investment column there were 20 missing data. During the preprocessing time we drop all missing data. This dataset was balanced using the SMOTE approach. SMOTE (Synthetic Minority Oversampling Strategy) is a statistical technique for evenly increasing the number of instances in your dataset. The component generates new instances based on existing minority situations you provide as input. The number of majority cases does not change as a result of SMOTE adoption.

D. Proposed Algrithms

The researchers used a publicly available dataset to evaluate seven machine learning techniques for detecting Insomnia. The specifics are as follows:

- Gradient Boosting
- Decision Tree
- K-Nearest Neighbor
- Logistic Regression
- Random Forest
- Ada Boost
- Support Vector Machine.

a) Gradient Boosting

Gradient boosting classifiers is a collection of machine learning techniques that combine numerous weaker models into a powerful, highly predictive output. Models of this kind are popular because of their ability to accurately categorize datasets. A block diagram of GB classifiers is shown in Figure 2.

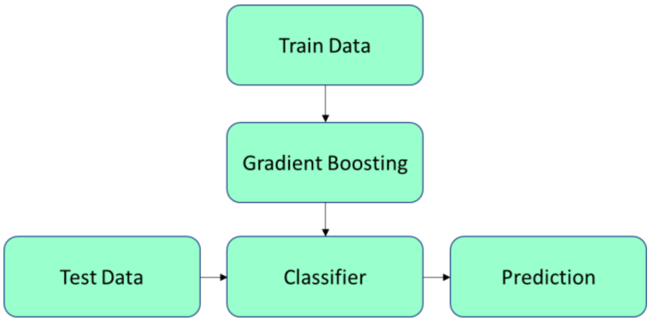

Figure 2: Diagram of Gradient boosting Classifier

GB builds an added substance model in a phase-by-stage way, taking into consideration the streamlining of any differentiable misfortune work. In each progression, the opposite slope of the binomial or multinomial aberrance misfortune work is used to fit n classes of relapse trees

b) Decision Tree

Decision tree approaches are often used in machine learning to tackle classification and regression issues. The two types of nodes formed by a root node are the internal node and the leaf node. Internal nodes are dubbed decision-makers because of their numerous branches, whereas leaf nodes are considered output nodes because they have no more branches. As a result, the tree's structure is altered. The decision tree classifier's fundamental architecture is shown in Figure 3.

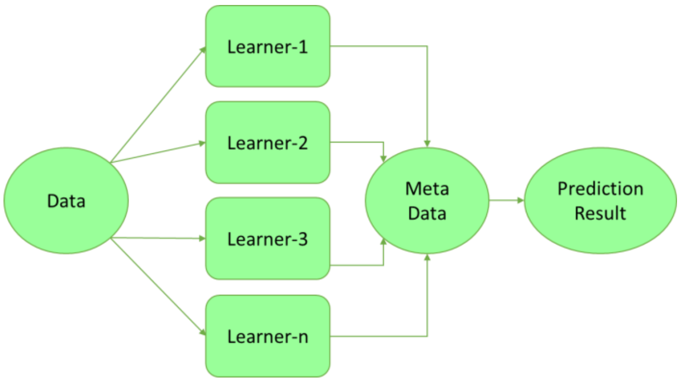

Figure 3: Diagram of Decision Tree Classifier

Data cleaning is the last thing on our minds when it comes to DT. It's easy to grasp since it's in the shape of a tree. The DT is simple to explain since it correlates to the processes that a person takes while making a real-life choice.

c) K-Nearest Neighbor

In a number of machine learning systems, supervised learning is applied. K-Nearest Neighbor is one of the simplest basic concepts. The KNN algorithm keeps track of all available data and classifies incoming data points based on how similar they are to data that has already been categorized. This implies that when fresh data is generated, the KNN technique may simply classify it into the proper category. Figure 4 illustrates the KNN diagram.

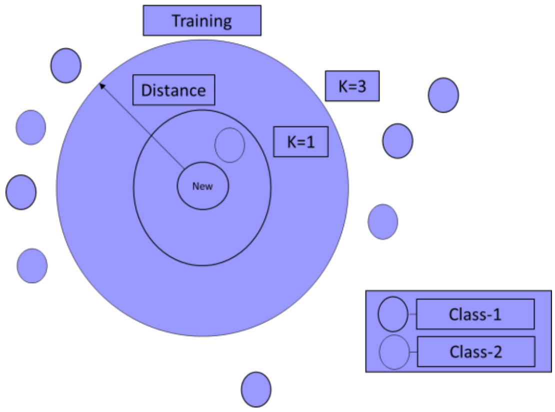

Figure 4: Diagram of K-Nearest Neighbor Classifier

The ease with which KNN can be implemented is one of the reasons it was chosen. It is more efficient when dealing with huge amounts of training data. It's a great starting point for dealing with noisy training data.

d) Logistic Regression

Despite its name, Logistic Regression (LR) is a machine learning method that is mostly utilized for binary classification problems. The LR may also be used for multiclass classification tasks that make use of one-vs.-rest learning methodologies. The block diagram of the logistic regression model is shown in Figure 5.

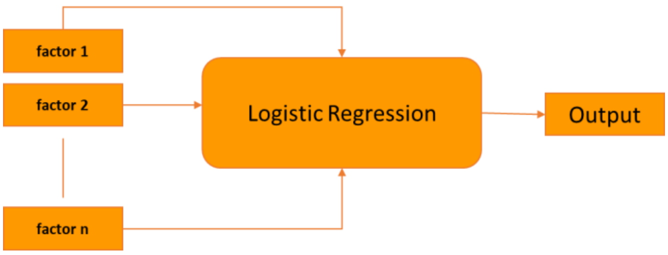

Figure 5: Diagram of Logistic Regression Classifier

The LR model is a linear machine learning model that employs the sigmoid function or its derivatives. The result of this operation is confined to the range [0, 1]. The probability of a certain class occurring is determined by how near the output is to 1.

e) Random Forest

As the initial algorithm, the Random Forest Classifier was used. It is made up of numerous independent trees, known as decision trees in RF, that are then trained using training sample data. The results of all of these trees are then put through a voting method in order to provide estimates. An RF classifier determines the final conclusion based on the majority of votes. A block diagram of RF classifiers is shown in Figure 6.

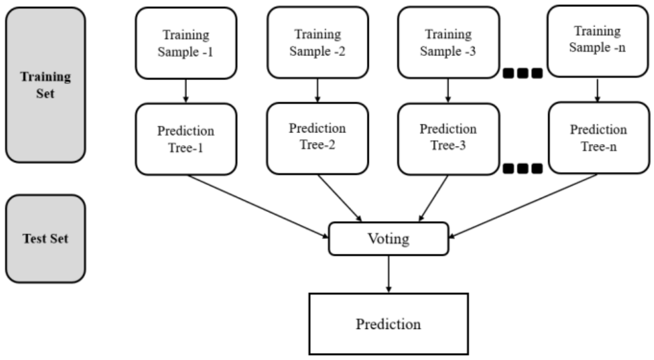

Figure 6: Diagram of Random Forest Classifier

Because RF takes less time to train than other algorithms, it will help us get better results. Because it delivers tremendous accuracy on bigger datasets, the size of the dataset has minimal influence on the accuracy. The lack of vast amounts of data has little impact on precision. It also works effectively since the conventional hyperparameters it employs transmit acceptable expectations in most cases. Because there are so few, it's critical to recognize the hyperparameters. Overfitting is the most common problem that develops in machine learning. In the random forest, though, it occurs less often. Having enough trees helps the classifier avoid overfitting.

f) Ada Boost Classifier

Boosting was invented in machine learning to investigate the possibility of transforming a collection of weak classifiers into a strong classifier. A suboptimal learner or classifier exceeds random guessing. This will be resistant to overfitting, as it will be composed of a large number of weak classifiers, each of which will perform better than random. Figure 7 shows the block diagram of Ada boost classifier.

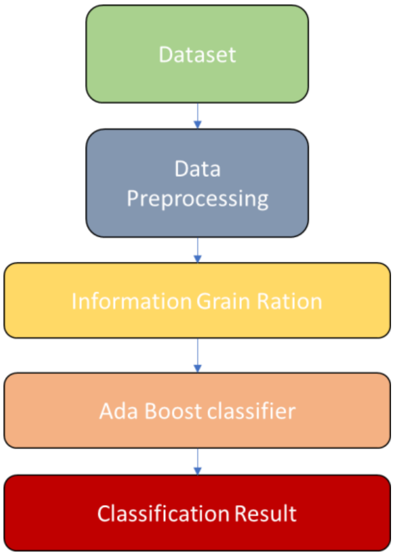

Figure 7: Flowchart of Decision Tree Classifier

Generally, a weak classifier is constructed using a simple threshold on a single feature. Positive characteristics are those that exceed the expected value; negative characteristics are those that exceed the projected value. AdaBoost is an acronym for 'Adaptive Boosting,' a strategy for converting weak learners or predictors to strong predictors in order to overcome classification challenges.

g) Support Vector classifier

The Support Vector Machine (SVM) is a versatile machine learning model that can do both regression and classification tasks. SVM is among one of the most well-known algorithms in ML research. The SVM model's goal is to split a given dataset into distinct classes in order to find the best hyperplanes. Figure 8 shows the support vector classifier's block diagram.

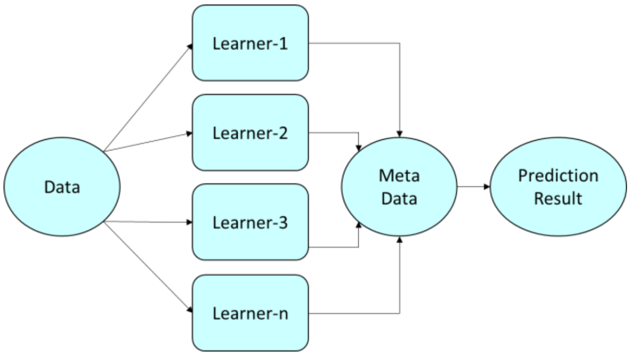

Figure 8: Diagram of SVC Classifier

Some of the advantages of using SVM include its ability to perform well with high-dimensional input fields. Furthermore, throughout the decision-making process, the SVM model allows for the selection of a number of kernel functions. One of its flaws is that it needs to be very carefully tuned in general, especially when the input dimension is bigger than the number of samples.

h) Evaluation Matrix

The evaluation matrix is a statistic that measures how well machine learning algorithms do when it comes to the confusion matrix. To evaluate the whole set of models, the confusion matrix will be employed. The confusion matrix depicts the frequency with which our models generate accurate and incorrect estimations.

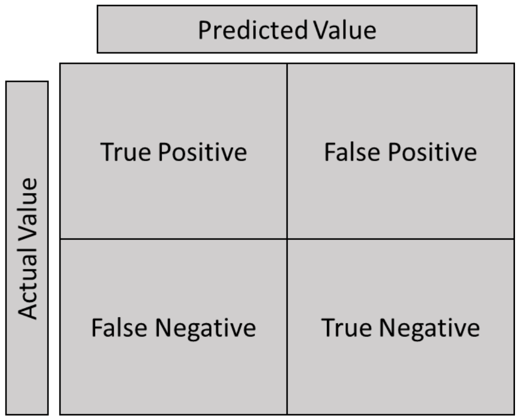

Figure 9: Block Diagram of Evaluation Matrix

Figure 9, false positives and negatives will be given to poorly protected values, whereas true positives and negatives will be assigned to successfully predicted values, as shown in Figure 9. After putting all the estimated values in the matrix together, the accuracy, precision-recall trade-off, and accuracy-recall trade-off were all looked at and calculated to determine how well the algorithm performed.

III. Data Analysis and Result

A. Relation Between Age and Insomnia

Age and Insomnia are explored in connection to one another. Figure 10 shows the relation between age and Insomnia.

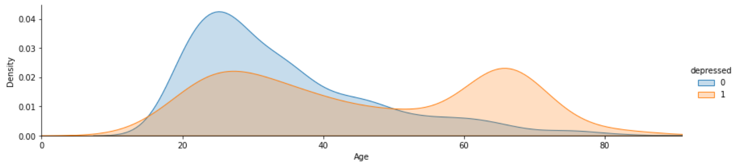

Figure 10: Relation between age and Insomnia

According to the graph above, Insomnia declines the most in middle age, between the ages of 45 and 50. Early adulthood Insomnia declines and late-life Insomnia increases are mostly the result of life-cycle gains and losses in marriage, job, and financial security. Adults 70 years of age and older are most susceptible to Insomnia because physical dysfunction and poor self-control exacerbate losses in identity and status.

B. Model Accuracy

a) Gradient Boosting

Figure 11 illustrates the Gradient boosting algorithm's classification accuracy.

```
[ ] y_pred_grad = grad_model.predict(X_test)
    print(classification_report(y_test, y_pred_grad))

              precision    recall  f1-score   support

           0       0.91      0.93      0.92        74
           1       0.91      0.88      0.89        57

    accuracy                           0.91       131
   macro avg       0.91      0.90      0.91       131
weighted avg       0.91      0.91      0.91       131
```

Figure 11: Classification accuracy of Gradient boosting classifier

The accuracy of the gradient boosting classifier was 91%. The healthy and Insomnia data prediction have an F1 score of 92% and 89%, respectively. Both have perfect precision. The gradient boosting algorithm's confusion matrix is shown in Figure 12.

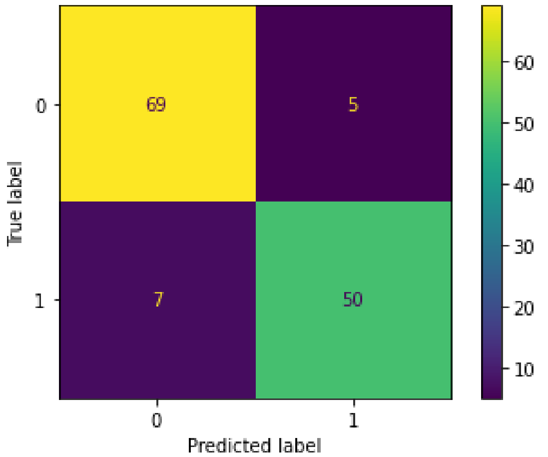

Figure 12: Confusion matrix of gradient boosting classifier

Figure 12 shows the forecast provided by the gradient boosting model. The anticipated outcome and the model's calculated performance are shown in the confusion matrix. 119 of the predictions were right, while just 12 were wrong. Figure 13 shows the ROC curve of the Gradient Boosting algorithm.

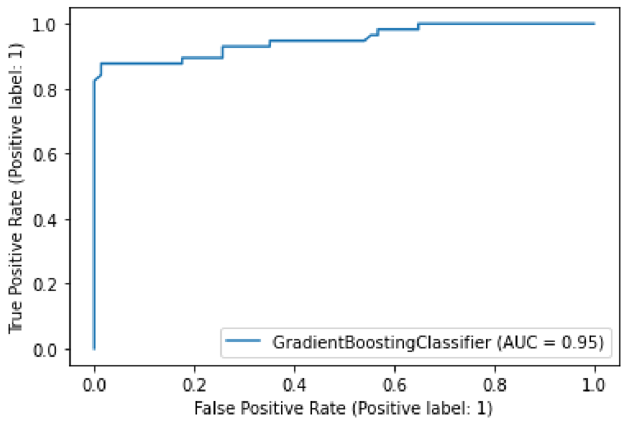

Figure 13: ROC curve of gradient boosting classifier

The receiver operating characteristic curve (ROC curve) is a graph that displays how well a classification model performs across all categorization levels. In this case the accuracy under the curve is 95%.

b) Decision Tree

Figure 14 depicts the classification accuracy of the decision tree method.

```
from sklearn import tree

dtree=tree.DecisionTreeClassifier()
dtree.fit(X_train,y_train)
pred_dtree=dtree.predict(X_test)

print(classification_report(y_test,pred_dtree))

              precision    recall  f1-score   support

           0       0.90      0.76      0.82        74
           1       0.74      0.89      0.81        57

    accuracy                           0.82       131
   macro avg       0.82      0.83      0.82       131
weighted avg       0.83      0.82      0.82       131
```

Figure 14: Classification accuracy of a decision tree classifier

The accuracy of the decision tree algorithm was 82 percent. The healthy and Insomnia data prediction has an F1 score of 82% and 81%, respectively. The decision tree algorithm's confusion matrix is shown in Figure 15.

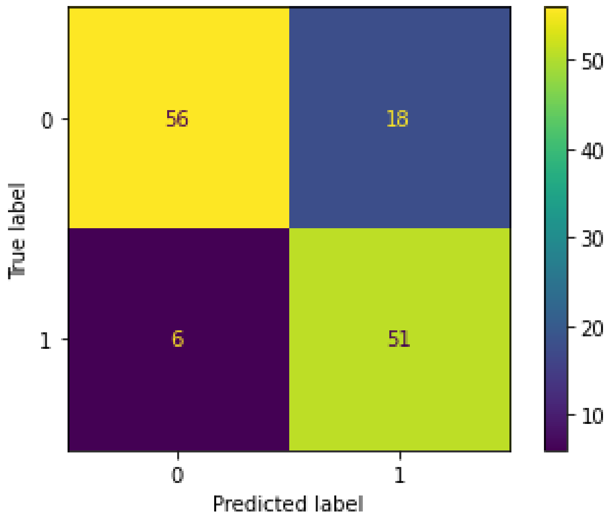

Figure 15: Confusion matrix of a decision tree classifier

Figure 15 shows the forecast provided by the decision tree model. The anticipated outcome and the model's calculated performance are shown in the confusion matrix. There were 107 accurate and 24 erroneous guesses. Figure 16 shows the ROC curve of decision tree classifier.

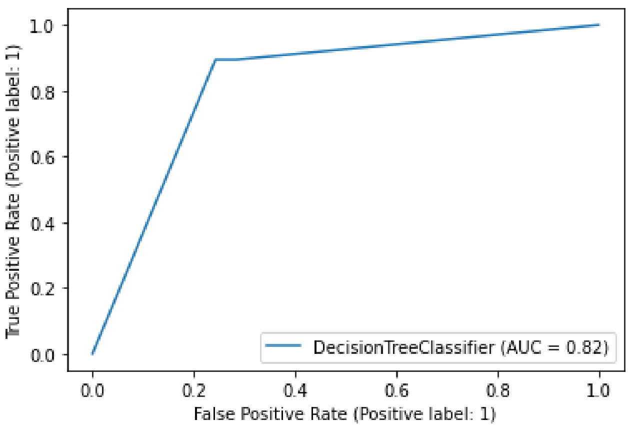

Figure 16: ROC curve of decision tree classifier

Decision tree classifier's outcome was lower than any other classifiers but higher than logistic regression classifier. In this case the accuracy under the curve was 82%.

c) KNN

Figure 17 illustrates the KNN algorithm's classification accuracy.

```
from sklearn.neighbors import KNeighborsClassifier
neigh = KNeighborsClassifier(n_neighbors=4)
neigh.fit(X_train, y_train)
ypred=neigh.predict(X_test)
print(classification_report(y_test,ypred))

              precision    recall  f1-score   support

           0       0.91      0.97      0.94        74
           1       0.96      0.88      0.92        57

    accuracy                           0.93       131
   macro avg       0.94      0.93      0.93       131
weighted avg       0.93      0.93      0.93       131
```

Figure 17: Classification accuracy of KNN classifier

The accuracy of the KNN algorithm was 93 percent. The healthy and Insomnia data prediction has an F1 score of 94% and 92%, respectively. The KNN algorithm's confusion matrix is shown in Figure 18.

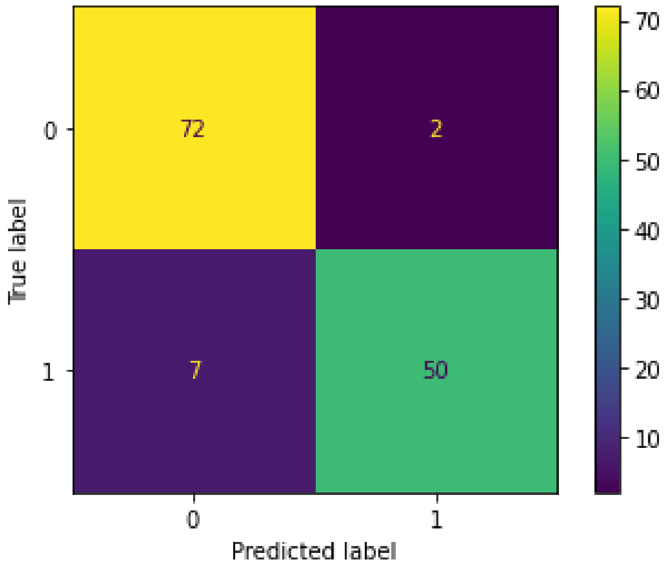

Figure 18: Confusion matrix of KNN classifier

Figure 18 shows the prediction provided by the KNN model. The anticipated outcome and the model's calculated performance are shown in the confusion matrix. 122 predictions were accurate, while 9 were wrong. Figure 19 shows the ROC curve of KNN classifier.

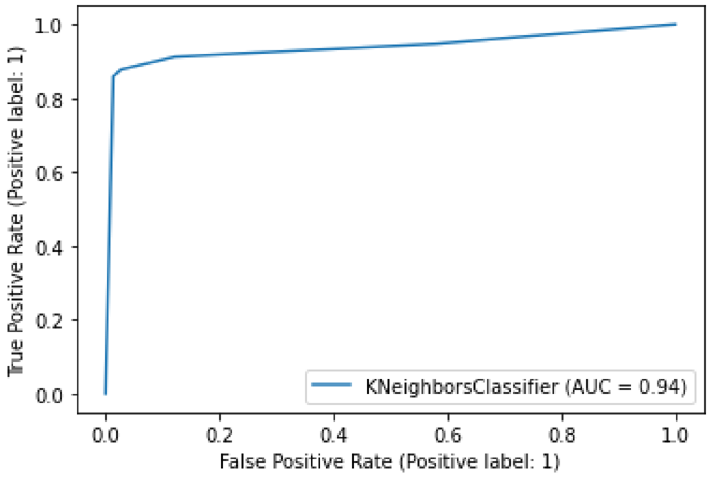

Figure 19: ROC curve of KNN classifier

The accuracy under the curve of KNN classifier was 94 percent. KNN classifier's outcome is more satisfied in this case.

d) Logistic Regression

The classification accuracy of the LR algorithm has been shown in figure 20.

```
[ ] from sklearn.linear_model import LogisticRegression

    lr = LogisticRegression(max_iter=2000)
    lr.fit(X_train, y_train)


    pred_lr=lr.predict(X_test)


    print(classification_report(y_test,pred_lr))

              precision    recall  f1-score   support

           0       0.74      1.00      0.85        74
           1       1.00      0.54      0.70        57

    accuracy                           0.80       131
   macro avg       0.87      0.77      0.78       131
weighted avg       0.85      0.80      0.79       131
```

Figure 20: Classification accuracy of LR classifier

The accuracy of the LR algorithm was 80 percent. For healthy and Insomnia data the f1-score was 85% and 70%, respectively. The LR algorithm's confusion matrix is shown in Figure 21.

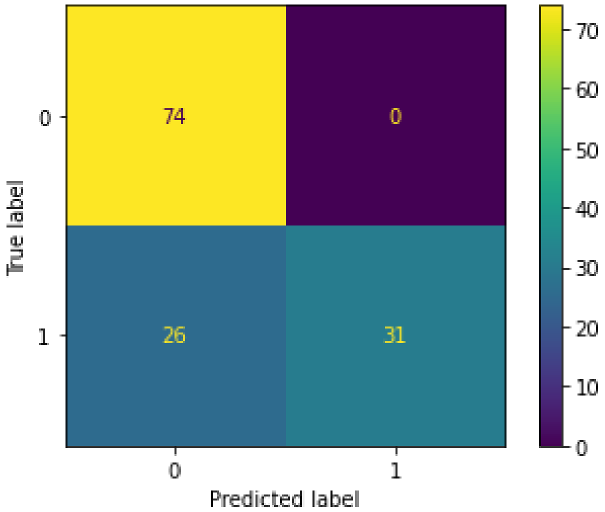

Figure 21: Confusion matrix of LR classifier

Figure 21 shows the prediction given by the LR model. The anticipated outcome and the model's calculated performance are shown in the confusion matrix. There were 105 accurate and 26 erroneous guesses. Figure 22 shows the ROC curve of the logistic regression classifier.

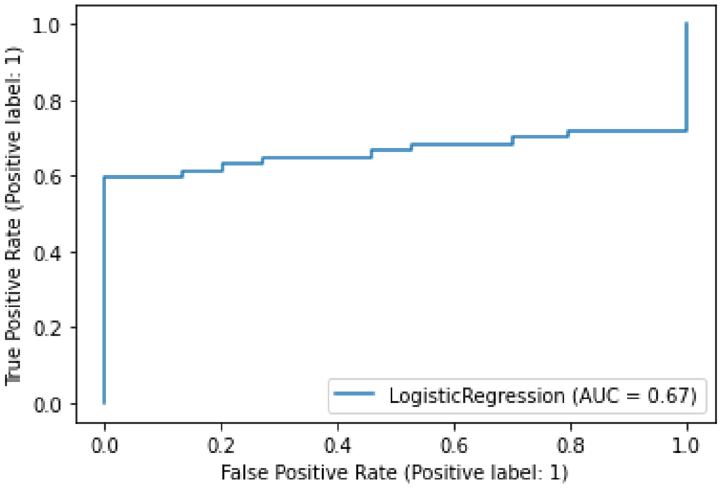

Figure 22: ROC curve of LR classifier

In this case, Logistic regression's outcome was the lowest among the seven-machine learning model. Also its accuracy under the curve was 67% which is lower than any other algorithm.

C. *Model Comperision*

The models are compared to those previously researched, as shown in Table 1. The Random Forest classifier beat all others in the system, as seen in the table.

Table 1: Performance Comparison

This Paper (Model Name)	Accuracy (%)	Reference Paper (Model Name)	Accuracy (%)
Decision Tree	82	Ref [29] Decision Tree	72.3
Logistic Regression	80	Ref [21] ANN	60.0
Gradient Boosting Classifier	91	Ref [30] XGBoost	92
K-Nearest Neighbor	93	Ref [32] KNN	72.10

Despite the fact that all of the methods in Table 1 are accurate, it is clear that only the random forest classifier approaches provide a considerable benefit in terms of accuracy. Using a Random Forest, KNN, SVM, and AB classifier, this work achieved 94, 93,93 and 93 percent accuracy, respectively. However, the accuracy of RF and SVM ref [29] and [31] was 76.6 percent and 90.1 percent, respectively. Using KNN, this work achieved 93 percent accuracy, whereas ref [32] achieved 72.10 percent accuracy using the same model.

IV. Conclusion

Today, Insomnia is recognized as a global super illness. In poor nations, around 75% of those who suffer from Insomnia remain untreated. This essay seeks to determine whether a person is depressed or not. Seven classifiers are utilized, including SVM, AB, GB, RF, DT, KNN, and LR, to get the best results. The outcome is noted without any filtering. In the future, a larger dataset, as well as machine learning models such as Extra Trees Classifier and Voting Classifier, may be utilized to enhance the framework models. Consequently, the consistency and appearance of the framework will be enhanced. The future of Insomnia detection systems will be defined by finding and using these tactics for detecting new and emerging threats. In return for merely supplying some basic information, the machine learning architecture may be able to assist the general public in determining the possibility of mental health related disease like Insomnia. The RF model has achieved the best accuracy metric combinations, which allows it to transform an incredibly nonlinear classification issue into a problem that can be solved linearly. This research might be seen as a first step in developing a comprehensive social media platform for diagnosing, forecasting, and providing treatments for users who are experiencing mental and psychological problems.

Data Availability Statement: The data used to support the findings of this study are freely available at https://ieee-dataport.org/documents/data-paper-are-under-medical-ethics-review-and-will-not-be-released-time-being

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors would like to confirm that there are no conflicts of interest regarding the study.

References

[1] A. D. Lopez, C. D. Mathers, M. Ezzati, D. T. Jamison, and C. J. Murray, “Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data,” Lancet, vol. 367, no. 9524, pp. 1747–1757, 2006.

[2] S. Moussavi, S. Chatterji, E. Verdes, A. Tandon, V. Patel, and B. Ustun, “Insomnia, chronic diseases, and decrements in health: results from the World Health Surveys,” Lancet, vol. 370, no. 9590, pp. 851–858, 2007.

[3] D. Lloyd-Jones, R. J. Adams, T. M. Brown et al., “Heart disease and stroke statistics-2010 update: a report from the american heart association,” Circulation, vol. 121, no. 7, pp. e46–e215, 2010.

[4] R. M. Carney and K. E. Freedland, “Insomnia in patients with coronary heart disease,” American Journal of Medicine, vol. 121, no. 11, supplement, pp. S20–S27, 2008.

[5] D. E. Bush, R. C. Ziegelstein, U. V. Patel et al., “Postmyocardial infarction Insomnia,” Evidence Report/Technology Assessment, no. 123, pp. 1–8, 2005.

[6] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, American Psychiatric Association, Washington, DC, USA, 4th edition, 1994.

[7] H. Dai, S. X. Zhang, K. H. Looi, R. Su, and J. Li, “Perception of health conditions and test availability as predictors of adults’ mental health during the covid-19 pandemic: A survey study of adults in Malaysia,” International Journal of Environmental Research and Public Health, vol. 17, no. 15, pp. 1–10, 2020, doi: 10.3390/ijerph17155498.

[8] N. Sahril, N. A. Ahmad, I. B. Idris, and R. Sooryanarayana, “Factors Associated with Mental Health Problems among Malaysian Children,” Children, no. 8, p. 119, 2021, doi: 10.3390/children8020119.

[9] J. Bilsen, “Suicide and Youth: Risk Factors,” Frontiers in Psychiatry, vol. 9, pp. 1–5, 2018, doi: 10.3389/fpsyt.2018.00540.

[10] L. Brådvik, “Suicide Risk and Mental Disorders,” International Journal of Environmental Research and Public Health, vol. 15, no. 9, 2018, doi: 10.3390/ijerph15092028.

[11] NHMS, “National Health and Morbidity Survey 2015. Vol. II: Non-Communicable Diseases, Risk Factors & Other Health Problems,” Ministry of Health, vol. 2. pp. 1–291, 2015.

[12] K. Katchapakirin, K. Wongpatikaseree, P. Yomaboot, and Y. Kaewpitakkun, “Facebook Social Media for Insomnia Detection in the Thai Community,” 2018, doi: 10.1109/JCSSE.2018.8457362.

[13] J. F. Greden and R. Garcia-tosi, Mental Health in the Workplace: Strategies and Tools to Optimize Outcomes. Springer, 2019.

[14] T. Kongsuk, S. Supanya, K. Kenbubpha, S. Phimtra, S. Sukhawaha, and J. Leejongpermpoon, “Services for Insomnia and suicide in Thailand,” WHO South-East Asia journal of public health, vol. 6, no. 1, pp. 34–38, 2017, doi: 10.4103/2224- 3151.206162.

[15] Y. Yang, C. Fairbairn, and J. F. Cohn, “Detecting Insomnia Severity from Vocal Prosody,” IEEE Transactions on Affective Computing, vol. 4, no. 2, pp. 142–150, 2013, doi: 10.1109/T-AFFC.2012.38.

[16] K. M. Griffiths, B. Carron-Arthur, A. Parsons, and R. Reid, “Effectiveness of Programs for Reducing The Stigma Associated with Mental Disorders. A Meta-Analysis of Randomized Controlled Trials,” World Psychiatry, vol. 13, no. 2, pp. 161–175, 2014, doi: 10.1002/wps.20129

[17] S. H. Yeoh, C. L. Tam, C. P. Wong, and G. Bonn, “Examining Depressive Symptoms and Their Predictors in Malaysia: Stress, Locus of Control, and Occupation,” Frontiers in Psychology, vol. 8, no. AUG, pp. 1–10, 2017, doi: 10.3389/fpsyg.2017.01411.

[18] G. Schomerus, H. Matschinger, and M. C. Angermeyer, “The Stigma of Psychiatric Treatment and Help-Seeking Intentions for Insomnia,” European Archives of Psychiatry and Clinical Neuroscience, vol. 259, no. 5, pp. 298–306, 2009, doi: 10.1007/s00406-009-0870-y.

[19] M.N Stolar, M. Lech, S.J. Stolar, N.B. Allen. Detection of adolescent Insomnia from speech using optimised spectral roll-off parameters. Biomedical Journal. 2018;2:10.

[20] R.S. Soundariya, M. Nivaashini, R.M. Tharsanee, P. Thangaraj, Application of Various Machine Learning Techniques in Sentiment Analysis for Insomnia Detection. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-10S, August 2019.

[21] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, H. Ohsaki. Recognizing Insomnia from twitter activity. In Proceedings of the 33rd annual ACM conference on human factors in computing systems 2015 Apr 18 (pp. 3187-3196).

[22] A. Haque, M. Guo, A.S. Miner, L. Fei-Fei. Measuring Insomnia symptom severity from spoken language and 3D facial expressions. arXiv preprint arXiv:1811.08592. 2018 Nov 21.

[23] M.M. Aldarwish, H.F. Ahmad. Predicting Insomnia levels using social media posts. In2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS) 2017 Mar 22 (pp. 277-280). IEEE.

[24] M.De. Choudhury, M. Gamon, S. Counts, E. Horvitz. Predicting Insomnia via social media. In Proceedings of the 7th International AAAI Conference on Social Media and Weblogs (ICWSM’13) (July 2013).

[25] N.P. Shetty, B. Muniyal, A. Anand, S. Kumar, S. Prabhu. Predicting Insomnia using deep learning and ensemble algorithms on raw twitter data. International Journal of Electrical & Computer Engineering (2088-8708). 2020 Aug 1;10.

[26] L. Cao, S. Guo, Z. Xue, Y. Hu, H. Liu, T.E. Mwansisya, W. Pu, B. Yang, C. Liu, J. Feng, E.Y. Chen. Aberrant functional connectivity for diagnosis of major depressive disorder: a discriminant analysis. Psychiatry and clinical neurosciences. 2014 Feb;68(2):110-9.

[27] S.C. Liao, C.T. Wu, H.C. Huang, W.T. Cheng, Y.H. Liu. Major Insomnia detection from EEG signals using kernel eigen-filter-bank common spatial patterns. Sensors. 2017 Jun;17(6):1385.

[28] Insomnia Dataset. Available on: https://ieee-dataport.org/documents/data-paper-are-under-medical-ethics-review-and-will-not-be-released-time-being

[29] Priya, Anu & Garg, Shruti & Tigga, Neha. (2020). Predicting Insomnia, Insomnia and Stress in Modern Life using Machine Learning Algorithms. Procedia Computer Science. 167. 1258-1267. 10.1016/j.procs.2020.03.442.

[30] A. Sharma, W.J.M.I. Verbeke. Improving Diagnosis of Insomnia with XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081). Front. Big Data 2020, 3, 15.

[31] K. Hilbert, U. Lueken, M. Muehlhan, K. Beesdo-Baum, Separating generalized Insomnia disorder from major Insomnia using clinical, hormonal, and structural MRI data: A multimodal machine learning study. Brain Behav. 2017, 7, e00633.

[32] A. Priya, S. Garg, N.P. Tigga, Predicting Insomnia, Insomnia and Stress in Modern Life using Machine Learning Algorithms. Procedia Comput. Sci. 2020, 167, 1258–1267